Claims 1-19 are pending. The Examiner is respectfully requested to reconsider and withdraw the outstanding rejections in view of the amendments and remarks contained herein.

REJECTION UNDER 35 U.S.C. §102

Claims 1-6 and 11-16 stand rejected under 35 U.S.C. §102(b) as being anticipated by Hsu et al. (EP 1072986 A2). This rejection is respectfully traversed.

Hsu et al. is generally directed toward extracting data from semi-structured text. In particular, the Examiner relies on Hsu et al. to teach generation of contextual rules from tokenized training text, and subsequent partitioning of input non-training text into tokens using the contextual rules. However, paragraphs 42-45 of Hsu et al. reveals that an input non-training text sequence is divided into tokens before it is sent to the information extractor, and that it is the extractor that then uses the contextual rules to extract data. Therefore, Hsu et al. does not use the contextual rules to partition the text string into tokens. Rather, Hsu et al. appears to use a structure of the text to partition the text into tokens, but is actually rather vague as to the tokenization technique employed. Therefore, Hsu et al. does not teach, suggest, or motivate segmenting an input stream into predefined tokens based on pattern information contained in a context record that has been generated in association with tokens of the input stream.

Applicant's claimed invention is generally directed toward a context-aware tokenizer. In particular, Applicant's claimed invention is generally directed toward segmenting an input stream into predefined tokens based on pattern information contained in a context record that has been generated in association with tokens of the

input stream. For example, independent claim 1 as originally filed recites, "at least one context automaton module that generates a context record associated with tokens of an input data stream; a tokenizing automaton module having a token automaton that partitions said input data stream into predefined tokens based on pattern information contained in said token automaton while simultaneously verifying contextual appropriateness based on said context record." Independent claim 11 recites similar subject matter. Thus, Hsu et al. do not teach all of the limitations of the independent claims.

Accordingly, Applicant respectfully requests the Examiner reconsider and withdraw the rejection of independent claims 1 and 11 under 35 U.S.C. § 102(b), along with rejection on these grounds of all claims dependent therefrom.

REJECTION UNDER 35 U.S.C. §103

Claims 7 and 17 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Hsu et al. (EP 1072986 A2) in view of Reps (ACM 1998). This rejection is respectfully traversed.

Hsu et al. is generally directed toward extracting data from semi-structured text. In particular, the Examiner relies on Hsu et al. to teach generation of contextual rules from tokenized training text, and subsequent partitioning of input non-training text into tokens using the contextual rules. However, paragraphs 42-45 of Hsu et al. reveals that an input non-training text sequence is divided into tokens before it is sent to the information extractor, and that it is the extractor that then uses the contextual rules to extract data. Therefore, Hsu et al. does not use the contextual rules to partition the text string into tokens. Rather, Hsu et al. appears to use a structure of the text to partition

the text into tokens, but is actually rather vague as to the tokenization technique employed. Therefore, Hsu et al. does not teach, suggest, or motivate segmenting an input stream into predefined tokens based on pattern information contained in a context record that has been generated in association with tokens of the input stream.

Reps is generally directed toward "maximal munch" tokenization in linear time. In particular, the Examiner relies on Reps to teach a linear time operating constraint. However, Reps does not teach, suggest, or motivate segmenting an input stream into predefined tokens based on pattern information contained in a context record that has been generated in association with tokens of the input stream.

Applicant's claimed invention is generally directed toward a context-aware tokenizer. In particular, Applicant's claimed invention is generally directed toward segmenting an input stream into predefined tokens based on pattern information contained in a context record that has been generated in association with tokens of the input stream. For example, independent claim 1 as originally filed recites, "at least one context automaton module that generates a context record associated with tokens of an input data stream; a tokenizing automaton module having a token automaton that partitions said input data stream into predefined tokens based on pattern information contained in said token automaton while simultaneously verifying contextual appropriateness based on said context record." Independent claim 11 recites similar subject matter. Thus, Hsu et al. and Reps do not teach, suggest, or motivate all of the limitations of the independent claims. These differences are significant.

Accordingly, Applicant respectfully requests the Examiner reconsider and withdraw the rejection of claims 7 and 17 under 35 U.S.C. § 103(a) in view of their dependence from allowable base claims 1 and 11.

Claims 8 and 18 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Hsu et al. (EP 1072986 A2) in view of Periera et al. (U.S. Pat. No. 5,781,884). This rejection is respectfully traversed.

Hsu et al. is generally directed toward extracting data from semi-structured text. In particular, the Examiner relies on Hsu et al. to teach generation of contextual rules from tokenized training text, and subsequent partitioning of input non-training text into tokens using the contextual rules. However, paragraphs 42-45 of Hsu et al. reveals that an input non-training text sequence is divided into tokens before it is sent to the information extractor, and that it is the extractor that then uses the contextual rules to extract data. Therefore, Hsu et al. does not use the contextual rules to partition the text string into tokens. Rather, Hsu et al. appears to use a structure of the text to partition the text into tokens, but is actually rather vague as to the tokenization technique employed. Therefore, Hsu et al. do not teach, suggest, or motivate segmenting an input stream into predefined tokens based on pattern information contained in a context record that has been generated in association with tokens of the input stream.

Periera et al. is generally directed toward grapheme to phoneme conversion of digit strings using weighted finite state transducers to apply grammar powers of a number basis. In particular, the Examiner relies on Periera et al. to teach a text to speech wherein the information from the partitioning influences the pronunciation of the text string. However, Periera et al. do not teach, suggest, or motivate segmenting an

input stream into predefined tokens based on pattern information contained in a context record that has been generated in association with tokens of the input stream.

Applicant's claimed invention is generally directed toward a context-aware tokenizer. In particular, Applicant's claimed invention is generally directed toward segmenting an input stream into predefined tokens based on pattern information contained in a context record that has been generated in association with tokens of the input stream. For example, independent claim 1 as originally filed recites, "at least one context automaton module that generates a context record associated with tokens of an input data stream; a tokenizing automaton module having a token automaton that partitions said input data stream into predefined tokens based on pattern information contained in said token automaton while simultaneously verifying contextual appropriateness based on said context record." Independent claim 11 recites similar subject matter. Thus, Hsu et al. and Periera et al. do not teach, suggest, or motivate all of the limitations of the independent claims. These differences are significant.

Accordingly, Applicant respectfully requests the Examiner reconsider and withdraw the rejection of claims 8 and 18 under 35 U.S.C. § 103(a) in view of their dependence from allowable base claims 1 and 11.

Claims 9, 10 and 19 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over Hsu et al. (EP 1072986 A2) in view of Corston-Oliver et al. (U.S. Pub. No. 20020138248). This rejection is respectfully traversed.

Hsu et al. is generally directed toward extracting data from semi-structured text. In particular, the Examiner relies on Hsu et al. to teach generation of contextual rules from tokenized training text, and subsequent partitioning of input non-training text into

tokens using the contextual rules. However, paragraphs 42-45 of Hsu et al. reveals that an input non-training text sequence is divided into tokens before it is sent to the information extractor, and that it is the extractor that then uses the contextual rules to extract data. Therefore, Hsu et al. does not use the contextual rules to partition the text string into tokens. Rather, Hsu et al. appears to use a structure of the text to partition the text into tokens, but is actually rather vague as to the tokenization technique employed. Therefore, Hsu et al. do not teach, suggest, or motivate segmenting an input stream into predefined tokens based on pattern information contained in a context record that has been generated in association with tokens of the input stream.

Corston-Oliver et al. is generally directed toward linguistically elegant text compression. In particular, the Examiner relies on Corston-Oliver to teach a message parser coupled to a linguistic analyzer, wherein an input message contains Japanese text that inherently lacks word space indicators. However, Corston-Oliver et al. do not teach, suggest, or motivate segmenting an input stream into predefined tokens based on pattern information contained in a context record that has been generated in association with tokens of the input stream.

Applicant's claimed invention is generally directed toward a context-aware tokenizer. In particular, Applicant's claimed invention is generally directed toward segmenting an input stream into predefined tokens based on pattern information contained in a context record that has been generated in association with tokens of the input stream. For example, independent claim 1 as originally filed recites, "at least one context automaton module that generates a context record associated with tokens of an input data stream; a tokenizing automaton module having a token automaton that

partitions said input data stream into predefined tokens based on pattern information contained in said token automaton while simultaneously verifying contextual appropriateness based on said context record." Independent claim 11 recites similar subject matter. Thus, Hsu et al. and Corston-Oliver et al. do not teach, suggest, or motivate all of the limitations of the independent claims. These differences are significant.

Accordingly, Applicant respectfully requests the Examiner reconsider and withdraw the rejection of claims 9, 10, and 19 under 35 U.S.C. § 103(a) in view of their dependence from allowable base claims 1 and 11.

CONCLUSION

It is believed that all of the stated grounds of rejection have been properly traversed, accommodated, or rendered moot. Applicant therefore respectfully requests that the Examiner reconsider and withdraw all presently outstanding rejections. It is believed that a full and complete response has been made to the outstanding Office Action, and as such, the present application is in condition for allowance. Thus, prompt and favorable consideration of this amendment is respectfully requested. If the Examiner believes that personal communication will expedite prosecution of this application, the Examiner is invited to telephone the undersigned at (248) 641-1600.

Respectfully submitted,

Dated: Dec 8, 2005

By: _____
Gregory A. Stobbs
Reg. No. 28,764

HARNESS, DICKEY & PIERCE, P.L.C.
P.O. Box 828
Bloomfield Hills, Michigan 48303
(248) 641-1600
GAS/JSB/kp